



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2020 June ; 29(6): 1204–1213. doi:10.1158/1055-9965.EPI-19-1262.

APOB genotypes and CDH13 haplotypes in the cholesterol-related pathway genes predict non-small cell lung cancer survival

Wei Deng^{1,2,3}, **Hongliang Liu**^{2,3}, **Sheng Luo**⁴, **Jeffrey Clarke**^{2,5}, **Carolyn Glass**^{2,6}, **Li Su**⁷, **Lijuan Lin**⁷, **David C. Christiani**^{7,8}, **Qingyi Wei**^{2,3,5}

¹Department of Experimental Research, Guangxi Medical University Cancer Hospital, Nanning, Guangxi 530021, China.

²Duke Cancer Institute, Duke University Medical Center, Durham, NC 27710, USA

³Department of Population Health Sciences, Duke University School of Medicine, Durham, NC 27710, USA

⁴Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA

⁵Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA

⁶Department of Pathology, Duke University School of Medicine, Durham, NC 27710, USA

⁷Departments of Environmental Health and Department of Epidemiology, Harvard School of Public Health, Boston, MA, 02115 USA

⁸Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA

Abstract

Background: Several oncogenic signals are involved in the synthesis, metabolism, transportation and modulation of cholesterol. However, the roles of genetic variants of the cholesterol pathway genes in cancer survival remain unclear.

Methods: We investigated associations between 26,781 common single-nucleotide polymorphisms (SNPs) in 209 genes of the cholesterol pathway and non-small cell lung cancer (NSCLC) survival by utilizing genotyping datasets from two published genome-wide association studies (GWASs). We used multivariate Cox proportional hazards regression and expression quantitative trait loci (eQTL) analyses to identify survival-associated SNPs and their correlations with the corresponding mRNA expression, respectively. We also used Kaplan-Meier survival

Corresponding Author: Qingyi Wei, Department of Population Health Sciences, Duke University School of Medicine, Duke Cancer Institute, Duke University Medical Center, GSRB1, Room3015C, MailBox103858, 595 LaSalle Street, Durham, NC 27710. Phone: 919-660-0562; qingyi.wei@duke.edu.

Disclosure of Potential Conflicts of Interest

Jeffrey M. Clarke declares the following conflict of interest: he has ongoing research projects funded by Bristol-Myers Squibb, Eli Lilly, Genentech, Spectrum, Adaptimmune, Medpacto, Bayer, AbbVie, Moderna, GlaxoSmithKline, Array, and AstraZeneca; he has been funded for his speeches by Merck; he has served on the Advisory Boards of AstraZeneca, Guardant, Merck, Eli Lilly, Pfizer, and NGM Biopharmaceuticals; his travels have been supported by Merck, AstraZeneca, Pfizer, NGM Bio; and he has served on Board of Director for Lung Cancer Initiative of North Carolina (uncompensated). Other authors declare no conflict of interest.

analysis and bioinformatics functional prediction to further evaluate the identified independent SNPs.

Results: We found five independent SNPs (*APOB* rs1801701C>T; *CDH13* rs35859010 C>T, rs1833970 T>A, rs254315 T>C and rs425904 T>C) to be significantly associated with NSCLC survival in both discovery and replication datasets. When the unfavorable genotype (*APOB* rs1801701CC) and haplotypes (*CDH13* rs35859010-rs1833970-rs254315-rs425904 C-A-T-C and T-T-T-T) were combined into a genetic score as the number of unfavorable genotypes/haplotypes (NUGH) in the multivariate analysis, an increased NUGH was associated with a worse survival ($P_{\text{trend}} < 0.0001$). In addition, both *APOB* rs1801701T<C and *CDH13* rs425904C<T were correlated with mRNA expression of the genes in normal lung tissues from the genotype-tissue expression (GTEx) project.

Conclusions: Genetic variants of *APOB* and *CDH13* in the cholesterol pathway were associated with NSCLC survival, possibly by affecting their gene expression.

Impact: Genetic variants of *APOB* and *CDH13* in the cholesterol pathway may provide new scientific insights into NSCLC prognosis.

Introduction

Lung cancer is one of the most common cancer types in the United States and remains the leading cause of cancer death. It is estimated that in 2019 there will be 228,150 new cases of lung cancer and 142,670 related deaths, accounting for 24% of all cancer deaths (1). The 5-year survival rate of lung cancer had been improved gradually up to 19.4% between 2009 and 2015 (2), but still strikingly lower than that of other cancers, especially compared with those that have similar morbidities (3). Therefore, additional research is needed to search for appropriate biomarkers for treatment response and thus survival of patients with lung cancer.

There are different histological subtypes of lung cancer, and non-small cell lung cancer (NSCLC) accounts for 85% of all lung cancer patients, which is further classified as lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) (4,5). NSCLC treatments include surgery, radiation and chemotherapy as well as the targeted therapy, depending upon histological typing and staging. It has been reported that cancer patients often adopt lifestyle changes after their diagnosis and treatments to improve their health status. For example, there was a three-fold increase in food supplemental use after patients were diagnosed with cancer (6), and approximately 70% of breast and prostate cancer survivors were overweight or obese after cancer diagnosis and successful treatment (7). Therefore, there is a growing concern that both nutritional status and an improper diet may have some significant effects on weakening the outcomes of treatment in cancer patients (8). In particular, dietary cholesterol as well as genes involved in its synthesis, metabolism, transportation and modulation have recently been a research focus of epidemiological, preclinical and clinical studies (9–13), which indicate that high dietary cholesterol intake may be prone to cancer, even influencing survival of cancer patients.

The controversial role of cholesterol in cancer development derived from several conflicting epidemiological studies that provided obscure results about associations between serum cholesterol levels and risk for certain cancer types (10,14,15), but preclinical studies more

consistently suggest a role of cholesterol in survival of cancer patients. For example, some studies provided evidence for a correlation between cholesterol synthesis and prognostic outcome, showing that several oncogenic signals, such as *PI3K/AKT/mTOR*, *RTK/RAS* and *TP53*, could modulate cholesterol synthesis in cancer cells (16–23). Studies in cultured cells and in animals also revealed that induction of cholesterol synthesis by the *AKT/mTORC1/SREBP* pathway contributed to cell growth (21) and promoted cancer aggressiveness and bone metastases (24,25). Furthermore, multiple cholesterol metabolites, such as steroids and oxysterols metabolized by mitochondrial cytochrome P450 family enzymes, were found to be involved in tumor growth and metastasis (26,27). Therefore, targeting the synthesis, transport, or metabolites of cholesterol may be alternative options for controlling cancer growth (28–33).

Although cholesterol effects differ by cancer type (34), the role of genetic variants in affecting the cholesterol pathway function and cancer survivals is not yet clear (16). Therefore, we conducted the present study to investigate the associations of genetic variants in the cholesterol pathway and related genes with NSCLC survival by using available genotyping datasets from two previously published genome-wide association studies (GWASs) of lung cancer.

Materials and Methods

Study populations

Two independent genotyping datasets were used for the discovery and replication, respectively, in the present study. Study populations of the discovery dataset derived from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial that was implemented between 1993 and 2001, in which baseline demographic characteristics and risk factors (such as smoking status) were archived, and whole blood samples were collected at enrollment (35–37). During the 13-year follow-up, there were totally 1,185 Caucasian participants who were confirmed to have developed NSCLC and whose histologic diagnosis, tumor stage, treatment method, survival time including overall survival (OS) and disease-special survival (DSS) time were recorded. Furthermore, their genomic DNA samples were extracted from the whole blood and genotyped with Illumina HumanHap240Sv1.0 and HumanHap550v3.0 (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1) (38–41).

The replication dataset included 984 Caucasian patients with histologically confirmed NSCLC from the Harvard Lung Cancer Susceptibility (HLCS) Study launched in 1992 (42). For these patients, demographic and clinical information along with survival time were collected. The whole blood samples were collected, and DNA was extracted with the Auto Pure Large Sample system for nucleic acid purification (QIAGEN Company, Venlo, Limburg, Netherlands) and genotyped by using the Illumina Humanhap610-Quad array. The genotyping data were used for imputation with the Mach3 software based on the sequencing data for Caucasians from the 1,000 Genomes Project.

The present study was approved by both the Internal Review Board of Duke University School of Medicine (#Pro00054575) and the dbGaP database administration (#6404). The comparison of the characteristics between the PLCO trial (n = 1,185) and the HLCS study (n

= 984) is presented in Supplementary Table 1. Because the discovery PLCO dataset had detailed genotyping data with more covariates available but the HLCS replication dataset did not, we could only use the PLCO dataset for further multivariate analyses after replication.

Gene and SNP selection

Based on the databases of the Molecular Signatures Databases (<http://software.broadinstitute.org/gsea/msigdb/index.jsp>), we included 209 genes related to the synthesis, metabolism and regulation of cholesterol in the analyses after excluding 60 duplicated genes, additional two genes withdrawn in NCBI and five genes on the X chromosome (Supplementary Table 2). We first extracted the genotyped data for these 209 candidate genes and their \pm 2-kb flanking regions from the PLCO dataset and then performed the imputation with IMPUTE2 using the sequencing data for Caucasians from the 1,000 Genomes Project database. As a result, a total of 26,781 SNPs (1,666 genotyped and 24,115 imputed), which met the criteria of a genotyping rate \geq 95%, a minor allelic frequency (MAF) \geq 5%, Hardy-Weinberg equilibrium (HWE) $\geq 1 \times 10^{-5}$ and an imputation info score \geq 0.8, were retained for subsequent analyses (Supplementary Figure 1B).

Multivariate Cox proportional hazards regression analysis

For the discovery PLCO dataset, we employed a single-locus analysis first with multivariate Cox proportional hazards regression models to assess the association between each of the 26,781 candidate SNPs and NSCLC survival by calculating hazards ratio (HR) and 95% confidence interval (CI). To correct for multiple testing, we first used false discovery rate (FDR) by a cutoff value of 0.2, followed by the Bayesian false discovery probability (BFDP) with a cutoff value of 0.8 as recommended for highly correlated SNPs as a result of imputation (43,44). We assigned a prior probability of 0.10 and a detectable upper boundary HR of 3.0 for an association with variant genotypes or minor alleles of the SNPs with $P < 0.05$. The multivariate Cox regression analysis and multiple test correction were performed by using the GenABEL package of R software (45).

For the HLCS replication dataset, associations between the identified significant SNPs and NSCLC survival in the PLCO dataset were further evaluated by using the multivariate Cox regression model with a significance level of $P < 0.05$. Finally, an inverse variance weighted meta-analysis was performed to combine the results of both discovery and replication datasets by PLINK 1.07.

Stepwise multivariate Cox regression analysis

To identify independent SNPs associated with NSCLC survival, we used a multivariate stepwise Cox model for the PLCO dataset, in which all the significant SNPs were included one by one, and their independence in predicting the outcome was evaluated by $P < 0.05$. In addition to the available demographic characteristics and clinical variables in the PLCO dataset, other 15 previously published SNPs (see the Results) associated with survival of NSCLC in the same PLCO dataset were also included in the model for further adjustment to confirm the newly identified independent survival-associated SNPs.

Combined effect analysis of all the independent SNPs

After the independent SNPs were confirmed, their effects on the NSCLC survival were assessed in the form of genotypes or haplotypes by multivariate analysis with adjustment for other covariates in the PLCO dataset in each of additive, dominant and recessive models. The genotype model with $HR > 1$ with $P < 0.05$ was regarded as the unfavorable genotype. If a cluster of independent SNPs was in a same gene, their haplotypes were also constructed and evaluated (46). For haplotype inference, we applied the HAPLOTYPE procedure of the SAS Genetics module given a multilocus sample of genetic marker genotypes under the assumption of Hardy-Weinberg equilibrium (HWE). The expectation-maximization (EM) algorithm was used to estimate the probability that each individual possesses a particular haplotype pair. The most likely haplotypes were used for each individual, and then the unfavorable haplotypes were identified by multivariate Cox regression analysis with $HR > 1$ and $P < 0.05$. Finally, a diplotype of each PLCO patient was assigned a value according to the number of unfavorable haplotypes on two strands of homologous chromosomes, and the association of diplotypes with OS and DSS of NSCLC were also evaluated using the same statistical method as for genotypes.

Once the unfavorable genotypes/haplotypes (UGH) were verified, they were combined into a number of unfavorable genotypes/haplotypes (NUGH) as a genetic score to assess the combined effect of all independent SNPs. We used Kaplan-Meier (K-M) curves and log-rank tests to evaluate the effects of NUGH on cumulative probability of OS and DSS with GraphPad Prism 8, and $P < 0.05$ was considered statistically significant.

Prediction model construction

We constructed a survival prediction model by using the receiver operating characteristic (ROC) curve with the “survival” and “timeROC” package of R software (version 3.5.0). Sensitivity, specificity and time-dependent area under the curve (AUC) were used to measure the ability of survival models to predict the NSCLC survival due to the effects of both clinical and genetic variables (47).

Stratified analysis

We performed a stratified analysis to evaluate associations between the UGH/NUGH and survival (both OS and DSS) of NSCLC in each stratum of the available covariates in the PLCO dataset, and the associations were assessed with a $P < 0.05$. We also assessed possible interactions with the Chi-square-based Q-test between genotypes/haplotypes and NSCLC among subgroups in the stratified analysis with $P_j < 0.05$.

Expression quantitative trait loci (eQTL) analysis

We performed the eQTL analysis to identify correlations between the independent SNPs and mRNA expression levels of the corresponding genes. Two approaches were used for the eQTL analysis; one was a linear regression model performed with the R software, in which the mRNA expression data were obtained from 373 European individuals in the 1,000 Genomes Project; and another was derived from two other GWAS datasets with normal lung tissue samples of 383 subjects and 369 whole blood samples, respectively, which are made available in the Genotype-Tissue Expression (GTEx) project (48,49).

Correlation between mRNA expression in lung cancer tissues and NSCLC survival

The correlations between mRNA expression levels of the SNP-associated genes and NSCLC survival were examined in 111 pairs of lung cancer (51 LUSC and 60 LUAD) tissues and their adjacent normal tissues from The Cancer Genome Atlas (TCGA) database by using a paired Student *t* test. We also used the Kaplan-Meier (K-M) survival curves to visualize the associations by an online tool (<http://kmplot.com/analysis/index.php?p=service&cancer=lung>), in which 1,926 NSCLC samples with published gene expression data and survival information from the caBIG, GEO and TCGA repositories were integrated; and *P* values for the K-M survival plot with HR and logrank were calculated, and the plots were made with R language (50).

All statistical analyses in the present study were performed by using the SAS software (version 9.4; SAS Institute, Cary, NC, USA), unless otherwise indicated.

Bioinformatics functional prediction

Bioinformatics functional prediction for each of the identified significant SNPs was performed with the online tools of SNPinfo (51) (<https://snpinfo.niehs.nih.gov>), RegulomeDB (52) (<http://www.regulomedb.org>) and HaploReg (53) (<https://pubs.broadinstitute.org/mammals/haploreg>).

Results

Basic characteristics of 1,185 NSCLC patients from the PLCO trial and 984 NSCLC patients from the HLCS study have been described elsewhere (54), and the detailed description of the present study is shown in Supplementary Figure 1A. Among the corresponding 26,781 SNPs of the 209 candidate genes in the cholesterol-related pathway, we identified 1,004 SNPs that were significantly associated with the OS in the PLCO dataset with multiple test correction by BFDP after failed by FDR, of which 24 SNPs remained significant after further replication by the HLCS dataset. As shown in Table 1 for the results of additive genetic models, these 24 SNPs are located in six genes, i.e., *APOB*, *ABCG5*, *RORA*, *CDH13*, *ABCG1* and *COMT*. Additional meta-analysis of the PLCO and HLCS datasets for these 24 identified SNPs showed the consistent results, and there was no heterogeneity between these two datasets (all $P_{het} > 0.05$).

Identification of independent SNPs among the 24 significant SNPs

After adjustment for other 15 previously reported survival-associated SNPs in the same PLCO dataset, five SNPs remained as independent survival predictors for further analysis. As shown in Table 2, *APOB* (rs1801701C>T) and *CDH13* (rs35859010C>T, rs1833970T>A, rs254315 T>C and rs425904T>C), as well as other demographic and clinical covariates except for radiotherapy, were independently associated with NSCLC survival (all $P < 0.05$).

The five independent SNPs are presented as marked in two separate Manhattan plots for PLCO and HLCS datasets (Supplementary Figure 1C and 1D). Furthermore, the regional association plot (<http://locuszoom.org/>) of each independent SNP is shown in

Supplementary Figure 2A–2E to illustrate their surrounding SNPs in the discovery dataset and the recombination rate estimated from HapMap Data Rel 22/phase II European population (55). Meanwhile, the pairwise linkage disequilibrium (LD) analysis using HaploView 4.2 software showed that four *CDH13* SNPs were in low LD (Supplementary Figure 2F).

Combined *APOB* genotypes and *CDH13* haplotypes and survival of NSCLC

For *APOB*, the rs1801701 CT+TT genotypes were associated with a better survival (HR = 0.73, 95% CI = 0.60–0.88, $P = 0.014$ for OS and HR = 0.74, 95% CI = 0.60–0.91, $P = 0.004$ for DSS), compared with the rs1801701 CC genotype. Therefore, the *APOB* rs1801701 CC genotype was the unfavorable genotype (Table 3).

For *CDH13*, considering the low LD of four SNPs in the same gene, their haplotypes were constructed and used in subsequent sequent analyses. The frequencies of *CDH13* haplotypes in the PLCO dataset were first estimated, and there were seven haplotypes named H1 – H7, of which haplotype H1 (C-T-T-T) was the most frequent (39.91%), followed by H2 (C-T-T-C, 18.34%), H3 (C-T-C-T, 11.11%), H4 (C-A-T-T, 10.04%), H5 (C-A-T-C, 9.83%), H6 (T-T-T-T, 6.89%) and H7 (T-T-C-T, 3.87%). Haplotypes H5 and H6 were found to be associated with worst NSCLC OS and DSS; when haplotype H5 and H6 were combined, they remained significantly associated with a worse NSCLC survival (HR = 1.38, 95% CI = 1.21–1.57, $P < 0.0001$ for OS and HR = 1.36, 95% CI = 1.19–1.56, $P < 0.0001$ for DSS (Table 3). As for *CDH13* diplotype, an increased number of unfavorable haplotypes on two strands of homologous chromosomes was associated with a worse survival in the multivariate analysis in the PLCO dataset (All $P < 0.05$ for OS and DSS, Supplementary Table 3).

To further evaluate the combined effect of these UGH on NSCLC OS and DSS in the PLCO dataset, we combined the significant unfavorable genotype (*APOB* rs1801701 CC) and diplotype (*CDH13* H5/H6) into a genetic score as the NUGH. As shown in Table 3, an increased NUGH was associated with a worse survival in the multivariate analysis in the PLCO dataset (both $P_{\text{trend}} < 0.0001$ for OS and DSS). We further used K-M survival curves to visualize these associations of the NUGH with NSCLC OS and DSS in Figure 1A–1D, in which NSCLC survival declined as the NUGH increased from 0, 1 to 2 UGH (Log-rank $P = 0.004$ for OS and Log-rank $P = 0.010$ for DSS). Similar results were observed, when the NUGH 2 group was compared with the NUGH 0–1 group (Log-rank $P = 0.007$ for OS and Log-rank $P = 0.012$ for DSS).

Combined *APOB* genotypes/*CDH13* haplotypes and survival prediction model

As shown in Figure 1, the ROC curves indicated an improved prediction performance with the addition of NUGH to the model with the covariates, compared to the model with the covariates only. We did not observe any difference in the prediction of 5-year survival by the UGH based on AUC and ROC curves. However, when we evaluated the 10-year NSCLC OS and DSS, the addition of UGH to the model with the covariates significantly increased the AUCs from 86.87% to 89.00% ($P < 0.0001$) and from 87.53% to 89.39% ($P = 0.0004$), respectively (Figure 1F and 1H). We further plotted the ROC curves for stage subgroups (i.e., I, II, III and IV) (Supplementary Figure 3), and there was no difference in AUCs by

stage. Although the 10-year NSCLC OS may not be clinically valuable, it may also reflect the roles of genetic factors in response to changes over time in lifestyle and dietary intake of cholesterol, which could influence the health status, even the outcome of clinical treatment of cancer patients (6,7).

Stratified analysis of associations between *APOB* genotypes/*CDH13* haplotypes and survival of NSCLC

As shown in Table 4, compared with those with 0–1 NUGH, individuals with 2 NUGH had a worse survival, consistently in each of the strata by all the covariates (all HR > 1.0 and $P < 0.05$), except for smoking status and chemotherapy. Meanwhile, the heterogeneity and interactions were also evaluated among these subgroups, and the results indicated that smoking status and chemotherapy had an interactive effect the associations of NUGH with OS ($P_{inter} = 0.026$ for smoking status and $P_{inter} = 0.0006$ for chemotherapy) and DSS ($P_{inter} = 0.050$ for smoking status and $P_{inter} = 0.002$ for chemotherapy) of NSCLC.

eQTL effects of *APOB* rs1801701 and *CDH13* rs425904 on mRNA expressions of their genes

For *APOB*, the eQTL analysis of the data from the GTEx project revealed that the rs1801701 T allele was significantly correlated with a lower expression level of *APOB* in 383 normal lung tissue samples ($P = 0.0269$; Figure 2A) but not in 369 whole blood samples. Because *APOB* expression data were not available in the 1000 Genomes Project, the eQTL analysis could not be performed for *APOB*.

For *CDH13*, the eQTL analysis of the data from the GTEx project revealed that the rs425904 C allele was significantly correlated with a lower expression level of *CDH13* in 383 normal lung tissue samples ($P = 0.0275$; Figure 2D) but not in 369 whole blood samples. In the RNA-Seq data of lymphoblastoid cell lines from the 1,000 Genomes Project, none of the four SNPs on *CDH13* (i.e., rs35859010, rs1833970, rs254315 and rs425904) showed a significant correlation with the mRNA expression of the gene in all three genetic models (Supplementary Figure 4A–4D); nor were the haplotypes of *CDH13* correlated with the mRNA expression levels (Supplementary Figure 4E).

Associations of mRNA levels of *APOB* and *CDH13* with survival of NSCLC

As shown in Figure 2B, in comparison with adjacent normal tissues, NSCLC (LUAD +LUSC) tumor tissues had a lower mRNA expression level of *APOB* ($P < 0.001$), which remained for both LUAD and LUSC samples, separately (all $P < 0.001$). Meanwhile, as shown in Figure 2C, a lower expression level of *APOB* was associated with a better survival of 1,926 NSCLC patients (HR = 1.28, 95% CI = 1.08–1.39 and Logrank $P = 0.0016$).

As shown in Figure 2E, in comparison with adjacent normal tissues, NSCLC (LUAD +LUSC) tumor tissues also had a lower mRNA expression level of *CDH13* ($P < 0.001$), which remained for LUAD samples ($P < 0.001$) but not for LUSC samples ($P = 0.073$). Meanwhile, as shown in Figure 2F, a higher expression level of *CDH13* was associated with a better survival of 1,926 NSCLC patients (HR = 0.76, 95% CI = 0.65–0.88 and Logrank $P = 0.00032$).

Bioinformatics functional prediction of the five independent SNPs

The results of functional prediction for the five independent SNPs identified as mentioned above revealed no evidence for functional relevance based on the SNPinfo, but there was some evidence for bioinformatics function based on RegulomeDB and HaploReg. Specifically, *APOB* rs1801701C>T and *CDH13* rs254315T>C are likely to have some effects on enhancer histone marks, DNase and motifs, while *CDH13* rs35859010C>T may have an effect on enhancer histone marks and motifs (Supplementary Table 4).

Discussion

In the current study, we performed a comprehensive analysis to investigate the associations between SNPs in genes involved in the cholesterol pathway and survival of NSCLC, utilizing two published GWAS datasets with a relatively long median follow-up time and strict quality control procedures. Five novel SNPs in two genes were identified, and the *APOB* genotypes and *CDH13* haplotypes were found to be associated with NSCLC survival.

APOB is located on chromosome 2p24.1 and encodes apolipoprotein B (ApoB), which is the main apolipoprotein of chylomicrons and low density lipoprotein (LDL) (56). LDL is commonly known as a “bad cholesterol” for both heart diseases and vascular diseases in general, while the functional roles of cholesterol as well as its carrier ApoB in cancer growth, especially in NSCLC, remain somewhat unclear. As mentioned above, increasing cellular cholesterol levels may promote proliferation and migration of cancer cells, likely leading to tumor progression (9–12). In one study that prospectively evaluated the associations between cancer mortality and circulating lipid biomarkers in 15,602 females, lipid levels were found to be associated with the total cancer deaths, including lung cancer (57). In addition, because a whole-exome sequencing study provided a strong evidence that *APOB* had an influence on LDL (58), we speculate that *APOB* may affect the progression of NSCLC through regulating the dietary intake and transport of cholesterol as well as the levels of downstream cholesterol metabolites. Therefore, reduction of the digestion and transport of cholesterol by *APOB* may be a potential adjuvant method for future NSCLC therapies, as a result of a better understanding of nutrient requirements, dietary intakes and nutrient metabolism in the patients.

CDH13 is located in chromosome 16q23.3 and encodes a member of the cadherin superfamily, which is localized on the surface of the cell membrane and is anchored by a glycosylphosphatidylinositol moiety, rather than by a transmembrane domain (59), affecting cellular behavior mainly through its signaling properties (60,61). Most of published studies focusing on the methylation of *CDH13* observed that the methylation level of *CDH13* was higher in NSCLC tumor tissues than in adjacent normal tissues and suggested that *CDH13* hypermethylation was associated with early recurrence and worse survival in NSCLC (62–64), although another study found that *CDH13* mRNA high expression levels were correlated with a better OS in adenocarcinoma patients (63). The mechanism underlying the observed association between down-regulation *CDH13* and poor prognosis of NSCLC may be related to the loss of *CDH13* ability to inhibit cell proliferation and invasiveness, which may increase susceptibility to apoptosis and reduce tumor growth (61).

On the other hand, *CDH13* is a putative receptor for a high molecular weight adiponectin, a cytokine produced by adipocytes, which also attracted research interest in recent years (65). There were several SNPs in *CDH13* that were reportedly to affect disease progression by influencing serum adiponectin levels (66,67), and the serum adiponectin level was found to be associated with prognosis of lung cancer (68). It is likely that the interaction between *CDH13* and adiponectin may be a potential signaling to influence NSCLC progression related to the cholesterol pathway. In the present study, we identified four novel SNPs and confirmed four haplotypes in *CDH13* to be associated with survival of NSCLC.

When we combined the unfavorable *APOB* genotype and *CDH13* haplotypes, we observed a dose-effect relationship between NUGH and both NSCLC OS and DSS. In addition, there was a weak interaction between smoking status and NUGH on survival of NSCLC patients. This may be related to the role of nicotine in tobacco that could affect serum cholesterol levels through *APOB*. In one study on the effect of nicotine on lipoprotein metabolism in rats, there was a significant increase in the levels of total cholesterol and ApoB in the sera of nicotine-treated rats (69). In the present study, we found that smoking status was strongly associated with a worse survival of NSCLC in the presence of the *APOB* rs1801701 CC genotype as well as *CDH13* haplotype H5 (C-A-T-C) and H6 (T-T-T-T) that affected the gene expression. Further stratified analysis found an interaction between chemotherapy and NUGH. Among the patients with 2 NUGH, those who did not receive chemotherapy tended to have a much worse NSCLC survival than those who received chemotherapy, but the difference was not statistically significant. This is likely due to selection bias that needs to be verified in future studies.

There are several limitations in the present study. First, different distributions of the demographic and clinic characteristics between the PLCO and HLCS study populations might have partially affected the validation; therefore, additional validation by other studies with more detailed prognostic factors, such as tumor stages and treatments, is needed to confirm these findings. Second, the sample sizes of two genotyping datasets were not large enough to allow for adequate subgroup analysis, particularly not for the FDR test, a more desired multiple test correction method. Third, because the present study used the genotyping data from populations of European ancestry, similar studies on other ethnic populations should be performed in the future. Fourth, we only analyzed associations between genetic variants in the identified genes in a selected pathway and survival, more survival-association studies should be called upon on genetic variants in other important biological pathway genes that are likely relevant to tumor phenotypes and treatment response in NSCLC patients. Finally, additional mechanistic studies should be performed to explore possible molecular mechanisms underlying the observed associations between the SNPs and survival of NSCLC patients.

In summary, the present study suggested a potential role of genetic variants of the cholesterol pathway genes *APOB* and *CDH13* in NSCLC survival, possibly through the modulation of the synthesis, transport and metabolism of cholesterol by these SNPs and genes, which may provide new scientific insights into NSCLC prognosis and clinical management, once replicated by other investigators.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the National Institute of Health (R01NS091307 to S. Luo); 2018 Guangxi One Thousand Young and Middle-Aged College and University Backbone Teachers Cultivation Program (GJR 2018-18 To W. Deng), the Duke Cancer institute as part of the P30 Cancer Center Support Grant (NIH/NCI CA014236 to M. Kastan); and the V Foundation for Cancer Research (D2017-19 to Q.Y. Wei). The Harvard Lung Cancer Susceptibility Study was supported by NIH grants R01CA092824 and R01CA074386 to D. C. Christiani.

The authors thank all the participants of the PLCO Cancer Screening Trial; Dongfang Tang, Sen Yang and Yuchen Zhao for their technical and statistical assistance; and the National Cancer Institute for providing the access to the data collected by the PLCO trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by the National Cancer Institute. The authors would also like to acknowledge the dbGaP repository for providing cancer genotyping datasets. The accession numbers for the datasets for lung cancer are phs000336.v1.p1 and phs000093.v2.p2. A list of contributing investigators and funding agencies for those studies can be found in the Supplemental Data.

Abbreviation Lists

<i>APOB</i>	apolipoprotein B
AUC	area under the receiver operating characteristic curve
BFDP	Bayesian false discovery probability
<i>CDH13</i>	cadherin 13
CI	confidence interval
DSS	disease-specific survival
EAF	effect allele frequency
eQTL	expression quantitative trait loci
GTE_x	genotype-tissue expression project
GWAS	Genome-Wide Association Study
HLCS	Harvard Lung Cancer Susceptibility
HR	hazards ratio
LD	linkage disequilibrium
LDL	low density lipoprotein
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
NSCLC	Non-small cell lung cancer
NUGH	number of unfavorable genotypes/haplotypes

OS	overall survival
PLCO	the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial
ROC	receiver operating characteristic
SNPs	single nucleotide polymorphisms
SREBP	sterol regulatory element binding protein
TCGA	the Cancer Genome Atlas
UGH	unfavorable genotypes/haplotype

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69:7–34. [PubMed: 30620402]
2. Surveillance The, Epidemiology, and End Results (SEER). Cancer Stat Facts: Lung and Bronchus Cancer; 2019 Available from: <https://seer.cancer.gov/statfacts/html/lungb.html>.
3. Surveillance The, Epidemiology, and End Results (SEER). Cancer Stat Facts: Common Cancer Sites; 2019 Available from: <https://seer.cancer.gov/statfacts/html/common.html>.
4. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Translational lung cancer research* 2016;5:288–300. [PubMed: 27413711]
5. Sher T, Dy GK, Adjei AA. Small cell lung cancer. *Mayo Clin Proc* 2008;83:355–67. [PubMed: 18316005]
6. Daenen LG, Cirkel GA, Houthuijzen JM, Gerrits J, Oosterom I, Roodhart JM, et al. Increased Plasma Levels of Chemoresistance-Inducing Fatty Acid 16:4(n-3) After Consumption of Fish and Fish Oil. *JAMA oncology* 2015;1:350–8. [PubMed: 26181186]
7. Demark-Wahnefried W, Aziz NM, Rowland JH, Pinto BM. Riding the crest of the teachable moment: promoting long-term health after the diagnosis of cancer. *J Clin Oncol* 2005;23:5814–30. [PubMed: 16043830]
8. Surwillo A, Wawrzyniak A. Nutritional assessment of selected patients with cancer. *Rocz Panstw Zakl Hig* 2013;64:225–33. [PubMed: 24325090]
9. Smith B, Land H. Anticancer activity of the cholesterol exporter ABCA1 gene. *Cell Rep* 2012;2:580–90. [PubMed: 22981231]
10. Allott EH, Howard LE, Cooperberg MR, Kane CJ, Aronson WJ, Terris MK, et al. Serum lipid profile and risk of prostate cancer recurrence: Results from the SEARCH database. *Cancer Epidemiol Biomarkers Prev* 2014;23:2349–56. [PubMed: 25304929]
11. Montero J, Morales A, Llacuna L, Lluís JM, Terrones O, Basanez G, et al. Mitochondrial cholesterol contributes to chemotherapy resistance in hepatocellular carcinoma. *Cancer Res* 2008;68:5246–56. [PubMed: 18593925]
12. Llaverias G, Danilo C, Mercier I, Daumer K, Capozza F, Williams TM, et al. Role of cholesterol in the development and progression of breast cancer. *Am J Pathol* 2011;178:402–12. [PubMed: 21224077]
13. Vassilev B, Sihto H, Li S, Holtta-Vuori M, Ilola J, Lundin J, et al. Elevated levels of StAR-related lipid transfer protein 3 alter cholesterol balance and adhesiveness of breast cancer cells: potential mechanisms contributing to progression of HER2-positive breast cancers. *Am J Pathol* 2015;185:987–1000. [PubMed: 25681734]
14. Ravnskov U, Rosch PJ, McCully KS. Statins do not protect against cancer: quite the opposite. *J Clin Oncol* 2015;33:810–1.
15. Nielsen SF, Nordestgaard BG, Bojesen SE. Statin use and reduced cancer-related mortality. *N Engl J Med* 2012;367:1792–802. [PubMed: 23134381]

16. Kuzu OF, Noory MA, Robertson GP. The Role of Cholesterol in Cancer. *Cancer Res* 2016;76:2063–70. [PubMed: 27197250]
17. Sui Z, Zhou J, Cheng Z, Lu P. Squalene epoxidase (SQLE) promotes the growth and migration of the hepatocellular carcinoma cells. *Tumour Biol* 2015;36:6173–9. [PubMed: 25787749]
18. Riganti C, Massaia M. Inhibition of the mevalonate pathway to override chemoresistance and promote the immunogenic demise of cancer cells: Killing two birds with one stone. *Oncoimmunology* 2013;2:e25770. [PubMed: 24327936]
19. Haskins JW, Zhang S, Means RE, Kelleher JK, Cline GW, Canfran-Duque A, et al. Neuregulin-activated ERBB4 induces the SREBP-2 cholesterol biosynthetic pathway and increases low-density lipoprotein uptake. *Science signaling* 2015;8:ra111. [PubMed: 26535009]
20. Gabitova L, Restifo D, Gorin A, Manocha K, Handorf E, Yang DH, et al. Endogenous Sterol Metabolites Regulate Growth of EGFR/KRAS-Dependent Tumors via LXR. *Cell Rep* 2015;12:1927–38. [PubMed: 26344763]
21. Porstmann T, Santos CR, Griffiths B, Cully M, Wu M, Leever S, et al. SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth. *Cell Metab* 2008;8:224–36. [PubMed: 18762023]
22. Dong F, Mo Z, Eid W, Courtney KC, Zha X. Akt inhibition promotes ABCA1-mediated cholesterol efflux to ApoA-I through suppressing mTORC1. *PLoS One* 2014;9:e113789. [PubMed: 25415591]
23. Ikonen E Cellular cholesterol trafficking and compartmentalization. *Nat Rev Mol Cell Biol* 2008;9:125–38. [PubMed: 18216769]
24. Yue S, Li J, Lee SY, Lee HJ, Shao T, Song B, et al. Cholesteryl ester accumulation induced by PTEN loss and PI3K/AKT activation underlies human prostate cancer aggressiveness. *Cell Metab* 2014;19:393–406. [PubMed: 24606897]
25. Thysell E, Surowiec I, Hornberg E, Crnalic S, Widmark A, Johansson AI, et al. Metabolomic characterization of human prostate cancer bone metastases reveals increased levels of cholesterol. *PLoS One* 2010;5:e14175. [PubMed: 21151972]
26. Fox EM, Andrade J, Shupnik MA. Novel actions of estrogen to promote proliferation: integration of cytoplasmic and nuclear pathways. *Steroids* 2009;74:622–7. [PubMed: 18996136]
27. McDonnell DP, Park S, Goulet MT, Jasper J, Wardell SE, Chang CY, et al. Obesity, cholesterol metabolism, and breast cancer pathogenesis. *Cancer Res* 2014;74:4976–82. [PubMed: 25060521]
28. Warita K, Warita T, Beckwith CH, Schurdak ME, Vazquez A, Wells A, et al. Statin-induced mevalonate pathway inhibition attenuates the growth of mesenchymal-like cancer cells that lack functional E-cadherin mediated cell cohesion. *Sci Rep* 2014;4:7593. [PubMed: 25534349]
29. Higgins MJ, Prowell TM, Blackford AL, Byrne C, Khouri NF, Slater SA, et al. A short-term biomarker modulation study of simvastatin in women at increased risk of a new breast cancer. *Breast Cancer Res Treat* 2012;131:915–24. [PubMed: 22076478]
30. Mo H, Elson CE. Studies of the isoprenoid-mediated inhibition of mevalonate synthesis applied to cancer chemotherapy and chemoprevention. *Exp Biol Med (Maywood)* 2004;229:567–85. [PubMed: 15229351]
31. Ginestier C, Monville F, Wicinski J, Cabaud O, Cervera N, Josselin E, et al. Mevalonate metabolism regulates Basal breast cancer stem cells and is a potential therapeutic target. *Stem Cells* 2012;30:1327–37. [PubMed: 22605458]
32. Kuzu OF, Gowda R, Sharma A, Robertson GP. Leelamine mediates cancer cell death through inhibition of intracellular cholesterol transport. *Mol Cancer Ther* 2014;13:1690–703. [PubMed: 24688051]
33. Murai T Cholesterol lowering: role in cancer prevention and treatment. *Biol Chem* 2015;396:1–11. [PubMed: 25205720]
34. Swinnen JV, Brusselmans K, Verhoeven G. Increased lipogenesis in cancer cells: new players, novel targets. *Curr Opin Clin Nutr Metab Care* 2006;9:358–65. [PubMed: 16778563]
35. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 2000;21:273s–309s. [PubMed: 11189684]

36. Hocking WG, Hu P, Oken MM, Winslow SD, Kvale PA, Prorok PC, et al. Lung cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *J Natl Cancer Inst* 2010;102:722–31. [PubMed: 20442215]
37. Oken MM, Marcus PM, Hu P, Beck TM, Hocking W, Kvale PA, et al. Baseline chest radiograph for lung cancer detection in the randomized Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *J Natl Cancer Inst* 2005;97:1832–9. [PubMed: 16368945]
38. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007;39:1181–6. [PubMed: 17898773]
39. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res* 2014;42:D975–9. [PubMed: 24297256]
40. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009;85:679–91. [PubMed: 19836008]
41. Landi MT, Consonni D, Rotunno M, Bergen AW, Goldstein AM, Lubin JH, et al. Environment And Genetics in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. *BMC Public Health* 2008;8:203. [PubMed: 18538025]
42. Zhai R, Yu X, Wei Y, Su L, Christiani DC. Smoking and smoking cessation in relation to the development of co-existing non-small cell lung cancer with chronic obstructive pulmonary disease. *Int J Cancer* 2014;134:961–70. [PubMed: 23921845]
43. Wakefield J A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* 2007;81:208–27. [PubMed: 17668372]
44. Györfy B, Györfy A, Tulassay Z. [The problem of multiple testing and solutions for genome-wide studies]. *Orv Hetil* 2005;146:559–63. [PubMed: 15853065]
45. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;23:1294–6. [PubMed: 17384015]
46. Stram DO. Multi-SNP Haplotype Analysis Methods for Association Analysis. *Methods Mol Biol* 2017;1666:485–504. [PubMed: 28980261]
47. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med* 2006;25:3474–86. [PubMed: 16220486]
48. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648–60. [PubMed: 25954001]
49. Lappalainen T, Sammeth M, Friedlander MR, Höfer PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;501:506–11. [PubMed: 24037378]
50. Györfy B, Surowiak P, Budczies J, Lanczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 2013;8:e82241. [PubMed: 24367507]
51. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res* 2009;37:W600–5. [PubMed: 19417063]
52. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–7. [PubMed: 22955989]
53. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 2016;44:D877–81. [PubMed: 26657631]
54. Wang Y, Liu H, Ready NE, Su L, Wei Y, Christiani DC, et al. Genetic variants in ABCG1 are associated with survival of nonsmall-cell lung cancer patients. *Int J Cancer* 2016;138:2592–601. [PubMed: 26757251]
55. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–7. [PubMed: 20634204]
56. Higuchi K, Hospattankar AV, Law SW, Meglin N, Cortright J, Brewer HB Jr. Human apolipoprotein B (apoB) mRNA: identification of two distinct apoB mRNAs, an mRNA with the

- apoB-100 sequence and an apoB mRNA containing a premature in-frame translational stop codon, in both liver and intestine. *Proc Natl Acad Sci U S A* 1988;85:1772–6. [PubMed: 2450346]
57. Chandler PD, Song Y, Lin J, Zhang S, Sesso HD, Mora S, et al. Lipid biomarkers and long-term risk of cancer in the Women’s Health Study. *Am J Clin Nutr* 2016;103:1397–407. [PubMed: 27099252]
58. Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, et al. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet* 2014;94:233–45. [PubMed: 24507775]
59. Vestal DJ, Ranscht B. Glycosyl phosphatidylinositol--anchored T-cadherin mediates calcium-dependent, homophilic cell adhesion. *J Cell Biol* 1992;119:451–61. [PubMed: 1400585]
60. Lee SW. H-cadherin, a novel cadherin with growth inhibitory functions and diminished expression in human breast cancer. *Nat Med* 1996;2:776–82. [PubMed: 8673923]
61. Andreeva AV, Kutuzov MA. Cadherin 13 in cancer. *Genes Chromosomes Cancer* 2010;49:775–90. [PubMed: 20607704]
62. Kim JS, Han J, Shim YM, Park J, Kim DH. Aberrant methylation of H-cadherin (CDH13) promoter is associated with tumor progression in primary nonsmall cell lung carcinoma. *Cancer* 2005;104:1825–33. [PubMed: 16177988]
63. Xue R, Yang C, Zhao F, Li D. Prognostic significance of CDH13 hypermethylation and mRNA in NSCLC. *Oncotargets Ther* 2014;7:1987–96. [PubMed: 25382980]
64. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, et al. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med* 2008;358:1118–28. [PubMed: 18337602]
65. Hug C, Wang J, Ahmad NS, Bogan JS, Tsao TS, Lodish HF. T-cadherin is a receptor for hexameric and high-molecular-weight forms of Acrp30/adiponectin. *Proc Natl Acad Sci U S A* 2004;101:10308–13. [PubMed: 15210937]
66. Morisaki H, Yamanaka I, Iwai N, Miyamoto Y, Kokubo Y, Okamura T, et al. CDH13 gene coding T-cadherin influences variations in plasma adiponectin levels in the Japanese population. *Hum Mutat* 2012;33:402–10. [PubMed: 22065538]
67. Putku M, Kals M, Inno R, Kasela S, Org E, Kozich V, et al. CDH13 promoter SNPs with pleiotropic effect on cardiometabolic parameters represent methylation QTLs. *Hum Genet* 2015;134:291–303. [PubMed: 25543204]
68. Kerenidi T, Lada M, Tsaroucha A, Georgoulas P, Mystridou P, Gourgoulanis KI. Clinical significance of serum adipokines levels in lung cancer. *Med Oncol* 2013;30:507. [PubMed: 23430445]
69. Ashakumary L, Vijayammal PL. Effect of nicotine on lipoprotein metabolism in rats. *Lipids* 1997;32:311–5. [PubMed: 9076668]

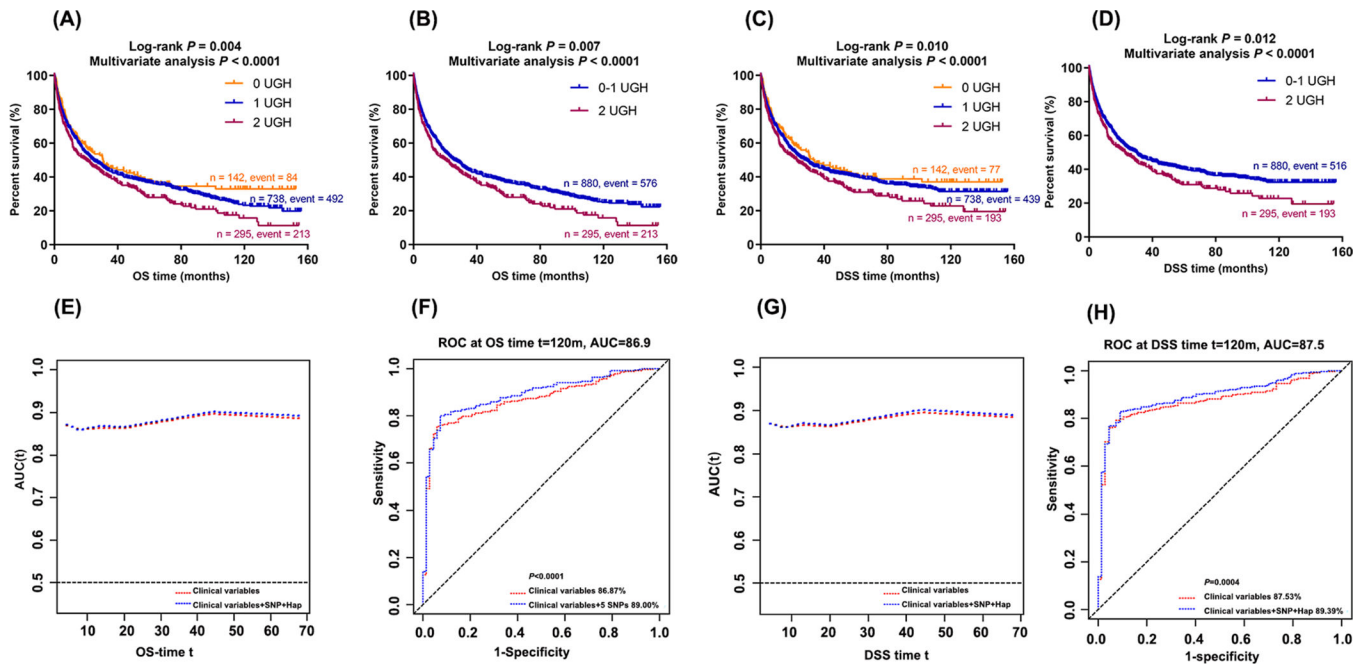


Figure 1. Association of combined *APOB* genotypes and *CDH13* haplotypes with NSCLC survival and prediction model constructed by ROC curve

Abbreviations: UGH, unfavorable genotypes (*APOB* rs1801701 CC) and haplotypes (*CDH13* H5 and *CDH13* H6); OS, overall survival; DSS, disease-special survival; ROC, receiver operating characteristic curve; AUC, Area under the curve. **(A)** Kaplan-Meier survival curves for the OS of combined UGH and **(B)** dichotomized groups by the number of UGH in the PLCO dataset; **(C)** Kaplan-Meier survival curves for the DSS of the combined UGH and **(D)** dichotomized groups by the number of UGH in the PLCO dataset. **(E)** OS and **(G)** DSS Time-dependent AUC estimation with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, surgery, principal component and the UGH; **(F)** Ten-year NSCLC OS and **(H)** Ten-year NSCLC DSS prediction by the ROC curve.

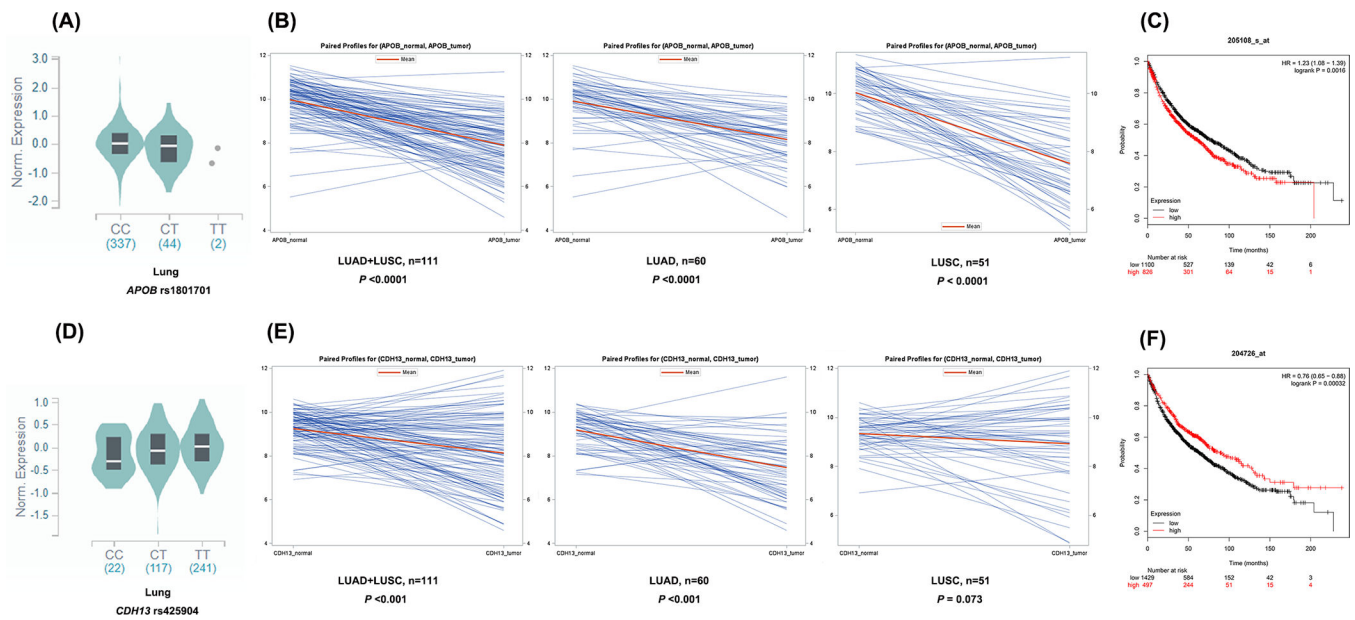


Figure 2. SNPs rs1801701 and rs425904 influenced the expressions of *APOB* and *CDH13* respectively and associated with the survival of NSCLC

Abbreviations: eQTL, expression quantitative trait loci analysis; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma. The eQTL analysis results for *APOB* rs1801701 (A) and *CDH13* rs425904 (D) in normal lung tissue in the GTEx Project. (B) Higher expression levels of *APOB* were found in the normal tissues, compared with both LUAD tissues and LUSC tissues. (E) Higher expression levels of *CDH13* were found in the normal tissues, compared with the LUAD tissues. (C) Higher *APOB* expression levels were associated with a better survival in NSCLC patients. (F) Higher *CDH13* expression levels were associated with a worse survival in NSCLC patients.

Table 1. Twenty-four SNPs consistently associated with survival of NSCLC patients in the PLCO and HLCS GWAS datasets

SNP	Allele ^a	Gene	PLCO (n=1185)				HLCS (n=984)				Combined-analysis			
			FDR	BFDP	EAF	HR (95% CI) ^b	P ^b	EAF	HR (95% CI) ^c	P ^c	P _{het} ^d	I ²	HR (95% CI) ^e	P ^e
rs1801701	C>T	<i>APOB</i>	0.27	0.34	0.09	0.75 (0.62–0.90)	0.002	0.09	0.81 (0.66–0.98)	0.032	0.592	0	0.78 (0.68–0.89)	0.0002
rs77105521	C>T	<i>ABCG5</i>	0.30	0.67	0.18	0.83 (0.73–0.95)	0.007	0.18	0.84 (0.72–0.97)	0.020	0.918	0	0.83 (0.76–0.92)	0.0003
rs76382861	G>A	<i>RORA</i>	0.29	0.47	0.15	0.81 (0.70–0.93)	0.002	0.13	0.81 (0.68–0.96)	0.017	0.998	0	0.81 (0.73–0.90)	0.0002
rs78023262	T>A	<i>RORA</i>	0.29	0.47	0.15	0.81 (0.70–0.93)	0.002	0.13	0.81 (0.68–0.96)	0.016	0.991	0	0.81 (0.73–0.90)	0.0002
rs35859010	C>T	<i>CDH13</i>	0.32	0.67	0.15	0.83 (0.72–0.95)	0.008	0.15	0.79 (0.65–0.96)	0.018	0.672	0	0.82 (0.73–0.91)	0.0004
rs1833970	T>A	<i>CDH13</i>	0.18	0.16	0.36	0.83 (0.74–0.92)	0.001	0.38	0.89 (0.79–1.00)	0.049	0.391	0	0.86 (0.79–0.93)	0.0001
rs2067861	C>T	<i>CDH13</i>	0.29	0.61	0.37	0.85 (0.77–0.95)	0.004	0.38	0.89 (0.79–1.00)	0.048	0.571	0	0.87 (0.80–0.94)	0.0003
rs9934609	C>T	<i>CDH13</i>	0.26	0.48	0.37	0.84 (0.75–0.94)	0.002	0.38	0.89 (0.79–1.00)	0.049	0.476	0	0.86 (0.80–0.94)	0.0004
rs9934700	C>T	<i>CDH13</i>	0.27	0.48	0.37	0.84 (0.75–0.94)	0.002	0.38	0.89 (0.79–0.99)	0.040	0.512	0	0.86 (0.80–0.93)	0.0003
rs6563943	G>A	<i>CDH13</i>	0.29	0.41	0.36	0.85 (0.76–0.94)	0.003	0.38	0.89 (0.79–1.00)	0.048	0.556	0	0.87 (0.80–0.94)	0.0004
rs72795378	G>A	<i>CDH13</i>	0.37	0.74	0.12	1.20 (1.04–1.38)	0.013	0.13	1.20 (1.03–1.40)	0.023	0.995	0	1.20 (1.08–1.33)	0.0006
rs17689520	C>G	<i>CDH13</i>	0.39	0.72	0.10	1.22 (1.04–1.42)	0.014	0.11	1.20 (1.00–1.42)	0.044	0.868	0	1.21 (1.08–1.36)	0.0014
rs60978336	C>A	<i>CDH13</i>	0.30	0.57	0.10	1.25 (1.06–1.46)	0.006	0.10	1.19 (1.00–1.42)	0.049	0.699	0	1.22 (1.09–1.38)	0.0008
rs12446784	A>G	<i>CDH13</i>	0.42	0.78	0.10	1.21 (1.03–1.41)	0.019	0.11	1.19 (1.00–1.42)	0.049	0.904	0	1.20 (1.07–1.35)	0.0020
rs72795399	A>G	<i>CDH13</i>	0.42	0.78	0.10	1.21 (1.03–1.41)	0.019	0.11	1.20 (1.00–1.43)	0.044	0.932	0	1.20 (1.07–1.35)	0.0019
rs11525693	C>G	<i>CDH13</i>	0.40	0.77	0.10	1.22 (1.04–1.43)	0.016	0.10	1.20 (1.01–1.43)	0.043	0.886	0	1.21 (1.08–1.36)	0.0014
rs254315	T>C	<i>CDH13</i>	0.29	0.67	0.15	1.21 (1.06–1.39)	0.006	0.15	1.19 (1.02–1.38)	0.029	0.844	0	1.20 (1.08–1.33)	0.0004
rs374476	A>G	<i>CDH13</i>	0.30	0.59	0.15	1.21 (1.06–1.38)	0.006	0.15	1.19 (1.02–1.39)	0.024	0.899	0	1.20 (1.09–1.33)	0.0003
rs425904	T>C	<i>CDH13</i>	0.41	0.78	0.20	1.16 (1.03–1.30)	0.017	0.21	1.15 (1.01–1.31)	0.039	0.917	0	1.16 (1.06–1.26)	0.0011
rs183436	A>C	<i>ABCG1</i>	0.31	0.72	0.32	0.86 (0.78–0.96)	0.007	0.31	0.88 (0.78–1.00)	0.046	0.734	0	0.87 (0.80–0.94)	0.0005
rs225390	G>A	<i>ABCG1</i>	0.29	0.72	0.33	0.86 (0.77–0.96)	0.006	0.33	0.87 (0.77–0.99)	0.029	0.854	0	0.87 (0.80–0.94)	0.0005
rs225395	C>T	<i>ABCG1</i>	0.29	0.61	0.34	0.85 (0.76–0.95)	0.003	0.34	0.88 (0.79–1.00)	0.043	0.637	0	0.87 (0.80–0.94)	0.0005
rs225398	C>G	<i>ABCG1</i>	0.29	0.56	0.34	0.86 (0.77–0.95)	0.005	0.35	0.88 (0.78–0.99)	0.038	0.768	0	0.87 (0.80–0.94)	0.0005
rs174699	T>C	<i>COMT</i>	0.18	0.13	0.05	1.45 (1.17–1.79)	0.001	0.05	1.30 (1.02–1.68)	0.037	0.527	0	1.39 (1.18–1.63)	0.0001

Abbreviations: SNP, single nucleotide polymorphism; NSCLC, non-small cell lung cancer; GWAS, genome-wide association study; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; EAF, effect allele frequency; HR, hazards ratio; CI, confidence interval; FDR, false discovery rate; BFDP, Bayesian false discovery probability; HLCS, Harvard Lung Cancer Susceptibility.

^aReference>effect allele.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

^bObtained from an additive genetic model with adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, and PC4;

^cObtained from an additive genetic model with adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, and PC3;

^d P_{het} : P -value for heterogeneity by Cochran's Q test;

^eMeta-analysis in the fixed-effects model.

Table 2. Five independent SNPs associated with survival of NSCLC patients in the PLCO Trial GWAS dataset

Variables	Category	Frequency	HR (95% CI) ^a	P ^a	HR (95% CI) ^b	P ^b
Age	Continuous	1185	1.03 (1.02–1.05)	<0.0001	1.04 (1.02–1.05)	<0.0001
Sex	Male/Female	698/487	0.76 (0.65–0.88)	0.0004	0.74 (0.63–0.87)	0.0002
Smoking status	Never/Current	115/423	1.73 (1.29–2.33)	0.0002	2.03 (1.5–2.74)	<0.0001
	Never/Former	115/647	1.79 (1.36–2.36)	<0.0001	2.01 (1.51–2.68)	<0.0001
Histology	Adeno/Squam	577/285	1.26 (1.04–1.51)	0.0173	1.25 (1.03–1.52)	0.0223
	Adeno/others	577/323	1.38 (1.16–1.64)	0.0003	1.40 (1.17–1.68)	0.0002
Tumor stage	I-III/IV	655/528	3.22 (2.65–3.91)	<0.0001	3.29 (2.70–4.02)	<0.0001
Chemotherapy	No/Yes	639/538	0.54 (0.45–0.65)	<0.0001	0.55 (0.46–0.66)	<0.0001
Radiotherapy	No/Yes	762/415	1.03 (0.87–1.21)	0.7578	0.99 (0.84–1.17)	0.9012
Surgery	No/Yes	637/540	0.21 (0.17–0.28)	<0.0001	0.19 (0.15–0.25)	<0.0001
<i>APOB</i> , rs1801701_T, C>T	CC/CT/TT	982/192/11	0.75 (0.62–0.90)	0.0024	0.79 (0.66–0.96)	0.0155
<i>CDHL3</i> , rs35859010_T, C>T	CC/CT/TT	856/297/32	0.85 (0.74–0.99)	0.0299	0.83 (0.71–0.96)	0.0108
<i>CDHL3</i> , rs1833970_A, T>A	TT/TA/AA	488/531/166	0.83 (0.75–0.93)	0.0011	0.83 (0.74–0.92)	0.0008
<i>CDHL3</i> , rs254315_C, T>C	TT/TC/CC	863/300/22	1.24 (1.08–1.42)	0.0023	1.33 (1.15–1.53)	<0.0001
<i>CDHL3</i> , rs425904_C, T>C	TT/TC/CC	757/371/57	1.23 (1.08–1.39)	0.0012	1.24 (1.09–1.40)	0.0007

Abbreviations: SNP; single nucleotide polymorphisms; NSCLC, non-small cell lung cancer; GWAS, genome-wide association study; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; HR, hazards ratio; CI, confidence interval.

^aStepwise analysis included age, sex, smoking status, tumor stage, histology, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, PC4.

^bFifteen published SNPs were used for post-stepwise adjustment: five SNPs were reported in previous publication (PMID: 27557513); One SNP was reported in the previous publication (PMID: 29978465); Two SNPs were reported in the previous publication (PMID: 30259978); Two SNPs were reported in the previous publication (PMID: 26757251); Three SNPs were reported in the previous publication (PMID: 30650190); Two SNPs were reported in the previous publication (PMID: 30989732); as well as adjusted by two significant SNPs have not published but been researching simultaneously in other pathway (related cell membrane transporters and trans-sulfation).

Table 3. Associations of combined *APOB* genotypes and *CDH13* haplotypes with survival of NSCLC patients

Genotypes/Haplotypes	Frequency ^d			OS ^b			DSS ^b		
	Deaths (%)	HR (95% CI)	P	Deaths (%)	HR (95% CI)	P	Deaths (%)	HR (95% CI)	P
<i>APOB</i>, rs1801701_T, C>T									
CC	663 (68.21)	1.00		594 (61.11)	1.00		594 (61.11)	1.00	
CT	119 (61.98)	0.72 (0.59–0.88)	0.016	108 (56.25)	0.73 (0.60–0.91)	0.004	108 (56.25)	0.73 (0.60–0.91)	0.004
TT	7 (63.64)	0.76 (0.36–1.61)	0.476	7 (63.64)	0.82 (0.39–1.73)	0.593	7 (63.64)	0.82 (0.39–1.73)	0.593
Trend test			0.002			0.006			0.006
Dominant									
CC ^c	663 (68.21)	1.00		594 (61.11)	1.00		594 (61.11)	1.00	
CT+TT	126 (62.07)	0.73 (0.60–0.88)	0.014	115 (56.65)	0.74 (0.60–0.91)	0.004	115 (56.65)	0.74 (0.60–0.91)	0.004
Recessive									
CC+CT	782 (67.18)	1.00		702 (60.31)	1.00		702 (60.31)	1.00	
TT	7 (63.64)	0.82 (0.39–1.72)	0.593	7 (63.64)	0.87 (0.41–1.84)	0.720	7 (63.64)	0.87 (0.41–1.84)	0.720
<i>CDH13</i> haplotypes									
H1	637 (67.91)	1.00		567 (60.45)	1.00		567 (60.45)	1.00	
H2	270 (62.65)	0.77 (0.67–0.89)	0.0004	245 (56.84)	0.80 (0.69–0.93)	0.003	245 (56.84)	0.80 (0.69–0.93)	0.003
H3	178 (68.20)	0.97 (0.82–1.15)	0.733	164 (62.84)	1.01 (0.84–1.20)	0.951	164 (62.84)	1.01 (0.84–1.20)	0.951
H4	145 (61.44)	0.78 (0.65–0.94)	0.008	127 (53.81)	0.79 (0.65–0.96)	0.019	127 (53.81)	0.79 (0.65–0.96)	0.019
H5	170 (73.59)	1.24 (1.05–1.47)	0.013	155 (67.10)	1.26 (1.05–1.51)	0.012	155 (67.10)	1.26 (1.05–1.51)	0.012
H6	117 (72.22)	1.27 (1.05–1.55)	0.017	105 (64.81)	1.25 (1.02–1.55)	0.034	105 (64.81)	1.25 (1.02–1.55)	0.034
H7	61 (67.03)	0.94 (0.72–1.23)	0.653	55 (60.44)	0.93 (0.71–1.24)	0.633	55 (60.44)	0.93 (0.71–1.24)	0.633
Combined <i>CDH13</i> haplotypes									
FH	1291 (65.97)	1.00		1158 (59.17)	1.00		1158 (59.17)	1.00	
UH ^d	287 (73.03)	1.38 (1.21–1.57)	<0.0001	260 (66.16)	1.36 (1.19–1.56)	<0.0001	260 (66.16)	1.36 (1.19–1.56)	<0.0001
NUGH^e									
0	84 (59.15)	1.00		77 (54.23)	1.00		77 (54.23)	1.00	
1	492 (66.67)	1.44 (1.13–1.82)	0.003	439 (59.49)	1.41 (1.10–1.81)	0.007	439 (59.49)	1.41 (1.10–1.81)	0.007
2	213 (72.20)	1.93 (1.49–2.50)	<0.0001	193 (65.42)	1.82 (1.43–2.46)	<0.0001	193 (65.42)	1.82 (1.43–2.46)	<0.0001

Genotypes/Haplotypes	Frequency ^a		OS ^b		DSS ^b		P
	Deaths (%)	HR (95% CI)	Deaths (%)	HR (95% CI)	Deaths (%)	HR (95% CI)	
Trend test							
NUGH							
0-1	880	1.00	567 (65.45)	1.00	516 (58.64)	1.00	<0.0001
2	295	1.42 (1.21-1.68)	213 (72.20)	1.42 (1.21-1.68)	193 (65.42)	1.41 (1.19-1.66)	<0.0001

Abbreviations: NSCLC, non-small cell lung cancer; HR, hazards ratio; CI, confidence interval; OS, overall survival; DSS, disease-specific survival; FH: favorable haplotypes; UH: unfavorable haplotypes; NUGH: number of unfavorable genotypes/haplotypes.

^aTen missing data were excluded.

^bAdjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, surgery and principal components.

^cUnfavorable genotype was *APOB* rs1801701 CC.

^dUnfavorable haplotypes were H5 (C-A-T-C) and H6 (T-T-T-T) for *CDH13* rs35859010-rs1833970-rs254315-rs425904.

^eNumber of unfavorable genotypes/haplotypes (NUGH) were assigned according to the *APOB* unfavorable genotype model and *CDH13* diplotype. 0 = no *APOB* rs1801701 CC or *CDH13* H5/H6; 1 = one *APOB* rs1801701 CC or *CDH13* H5/H6; 2 = both *APOB* rs1801701 CC and *CDH13* H5/H6.

Table 4. Stratified analysis for associations of *APOB* genotypes and *CDH13* haplotypes with survival of NSCLC patients

Characteristics	0-1 NUHG ^d		2 NUHG ^d		Multivariate analysis for OS ^b			Multivariate analysis for DSS ^b			P _{inter} ^c	
	All	OS	DSS	All	OS	DSS	HR (95% CI)	P	HR (95% CI)	P		
Age (years)												
71	479	290	258	155	109	101	1.27 (1.01-1.59)	0.039	1.26 (1.00-1.60)	0.051	0.321	
> 71	401	286	258	140	104	92	1.52 (1.20-1.91)	0.0004	1.50 (1.18-1.92)	0.001	0.321	
Sex												
Male	526	381	336	169	124	108	1.28 (1.04-1.57)	0.019	1.25 (1.01-1.56)	0.044	0.348	
Female	354	195	180	126	89	85	1.74 (1.34-2.25)	<0.0001	1.75 (1.34-2.29)	<0.0001	0.348	
Smoking status												
Never	86	46	45	28	16	15	0.92 (0.50-1.72)	0.802	0.88 (0.47-1.66)	0.694	0.050	
Current	311	198	171	106	68	64	1.18 (0.88-1.57)	0.266	1.27 (0.94-1.72)	0.116	0.050	
Former	483	332	300	161	129	114	1.62 (1.32-1.99)	<0.0001	1.56 (1.25-1.94)	<0.0001	0.050	
Histology												
Adeno	432	249	226	143	97	95	1.56 (1.23-1.99)	0.0003	1.63 (1.27-2.08)	0.0001	0.472	
Squamous	216	146	127	68	45	34	1.07 (0.76-1.51)	0.693	0.86 (0.58-1.27)	0.439	0.472	
Others	232	181	163	84	71	64	1.63 (1.22-2.16)	0.0009	1.66 (1.23-2.25)	0.001	0.472	
Tumor stage												
I-III	491	224	179	163	90	74	1.37 (1.06-1.76)	0.015	1.36 (1.03-1.79)	0.033	0.357	
IV	389	352	337	132	123	119	1.39 (1.12-1.72)	0.002	1.39 (1.12-1.73)	0.003	0.357	
Chemotherapy												
No	478	260	217	160	106	90	1.72 (1.36-2.17)	<0.0001	1.65 (1.28-2.12)	0.0001	0.002	
Yes	402	316	299	135	107	103	1.14 (0.91-1.42)	0.254	1.15 (0.92-1.45)	0.222	0.002	
Radiotherapy												
No	567	320	275	194	129	116	1.43 (1.16-1.76)	0.0007	1.47 (1.18-1.83)	0.0007	0.276	
Yes	313	256	241	101	84	77	1.35 (1.05-1.75)	0.019	1.29 (0.99-1.68)	0.058	0.276	
Surgery												
No	474	424	401	161	141	133	1.23 (1.01-1.49)	0.038	1.21 (0.99-1.48)	0.059	0.092	
Yes	406	152	115	134	72	60	1.78 (1.33-2.37)	<0.0001	1.89 (1.37-2.60)	<0.0001	0.092	

Abbreviations: OS, overall survival; DSS, disease-specific survival; NSCLC, non-small cell lung cancer; HR, hazards ratio; CI, confidence interval; NUGH, number of unfavorable genotypes/haplotypes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- ^aTen missing data were excluded;
- ^bAdjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, and PC4;
- ^c*P*inter: *P* value for interaction analysis between characteristic and UHG.